Appendix 1: Replacement Paragraphs of the Specification

Paragraph beginning on page 6 at line 6:

Now referring to FIGURE 4, a block diagram illustrates a third preferred embodiment of the system for determining a title from a document image according to the current invention. An image input unit 121 inputs a document image, and a document image storage unit 122 stores the inputted document image. A character row area determination unit 123 determines areas or minimal circumscribing rectangles that contain characters. The character row area determination unit 123 outputs the coordinates as well as the size of character row areas to a character recognition unit 124 as well as a title evaluation point determination unit 128. The character recognition unit 124 recognizes characters from character image portions in the character row areas. A reference describing character recognition, U.S. Pat. No. 5,966,464, is incorporated by reference herein in its entirety. The character recognition unit 124 generates corresponding character codes as well as other associated information. Other associated information includes the character recognition assurance level, the coordinates of a minimal circumscribing rectangle and the size of the rectangle. The outputs from the character recognition unit 124 are sent to a font determination unit 125, the title evaluation point determination unit 128, a natural language analysis unit 126 and a recognition result storage unit 129. The font determination unit 125 determines a font type and other associated information for each character and outputs the font information to the title evaluation point determination unit 128. A reference describing font determination, Japanese Patent Laid Publication Hei 9-319830, is incorporated by reference herein in its entirety. The natural language analysis unit 126 compares the recognized characters against a predetermined dictionary and determines whether or not the recognized characters match or resemble any of the predetermined titles or words in a dictionary. For example, the dictionary contains a set of predetermined suffixes which indicate a noun form and its corresponding statistical information. The natural language analysis unit 126 also outputs the determination information to the title evaluation point determination unit 128. A characteristics extraction unit 127 extracts information on certain layouts such as underlining, centering and the minimal circumscribing rectangle size from the input image and outputs the information to the title evaluation point determination unit 128. For example, if the character size is beyond 18-point in an A4 image, the minimal circumscribing rectangle containing the characters is assigned a high score. Similarly, a high score is assigned to a minimal circumscribing rectangle if a

number of characters or words in the rectangle is less than a predetermined number. For example, for the Japanese language, the predetermined number of characters may be set to twelve. The above and other predetermined numbers are user-definable.

Paragraph beginning on page 9 at line 29:

FIGURE 7 illustrates other acts involved in determining the likelihood based upon a number of characters according to the current invention. In act A401, a document image is inputted, and character row areas are determined in act A402. After the character image in the character row areas is converted into character codes, a number of characters is determined. The number of characters is compared to a predetermined threshold value in act A404. A set of predetermined threshold values is optionally stored in a statistical dictionary for different types of documents. If the number of characters is below the predetermined threshold value in act A405, a predetermined number of points is added to the likelihood for the character row area and a title area selection is determined based upon the total number of points in act A406. On the other hand, if the number of characters is above the predetermined threshold value in act A405, other predetermined processing is performed.